



IFWO

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/798,278

DATE: 08/31/2004

TIME: 12:50:47

Input Set : N:\Crf3\RULE60\10798278.raw
 Output Set: N:\CRF4\08312004\J798278.raw

1 <110> APPLICANT: ARAKI, HIROYUKI
 2 ENDO, KEIJI
 3 HAGIHARA, HIROSHI
 4 IGARASHI, KAZUAKI
 5 HAYASHI, YASUHIRO
 6 OZAKI, KATSUYA
 7 <120> TITLE OF INVENTION: HIGHLY PRODUCTIVE ALPHA-AMYLASES
 8 <130> FILE REFERENCE: 214377US0
 9 <140> CURRENT APPLICATION NUMBER: US/10/798,278
 10 <141> CURRENT FILING DATE: 2004-03-12
 11 <150> PRIOR APPLICATION NUMBER: US/09/971,611
 12 <151> PRIOR FILING DATE: 2001-10-09
 13 <150> PRIOR APPLICATION NUMBER: JP 2000/310605
 14 <151> PRIOR FILING DATE: 2001-10-11
 15 <160> NUMBER OF SEQ ID NOS: 51
 16 <170> SOFTWARE: PatentIn version 3.1
 18 <210> SEQ ID NO: 1
 19 <211> LENGTH: 1786
 20 <212> TYPE: DNA
 21 <213> ORGANISM: Bacillus sp. KSM-AP1378
 22 <220> FEATURE:
 23 <221> NAME/KEY: sig_peptide
 24 <222> LOCATION: (155)..(247)
 25 <223> OTHER INFORMATION:
 26 <221> NAME/KEY: mat_peptide
 27 <222> LOCATION: (248)..()
 28 <223> OTHER INFORMATION:
 --> 29 <221> CDS
 30 <222> LOCATION: (155)..(1702)
 31 <223> OTHER INFORMATION:
 --> 32 <400> 1
 33 cagcgtgata atataaattt gaaatgaaca cctataaaaa tatggtagcg attgcgcac 60
 34 gagaaaaaac ttgggagttt ggaagtgata ttaaaggatt tttttgact tggttgaaa 120
 35 acgcttgcatt aaattgaagg agagggtgtt tttt atg aaa ctt cat aac cgt ata 175
 Met Lys Leu His Asn Arg Ile
 36 -30 -25
 38 att agc gta cta tta aca cta ttg tta gct gta gct gtt ttg ttt cca 223
 39 Ile Ser Val Leu Leu Thr Leu Leu Ala Val Ala Val Leu Phe Pro
 40 -20 -15 -10
 41 tat atg acg gaa cca gca caa gcc cat cat aat ggg acg aat ggg acc 271
 42 Tyr Met Thr Glu Pro Ala Gln Ala His His Asn Gly Thr Asn Gly Thr
 43 -5 -1 1 5
 44 atg atg cag tat ttt gaa tgg cat ttg cca aat gac ggg aac cac tgg 319

ENTERED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/798,278

DATE: 08/31/2004

TIME: 12:50:47

Input Set : N:\Crf3\RULE60\10798278.raw

Output Set: N:\CRF4\08312004\J798278.raw

45	Met Met Gln Tyr Phe Glu Trp His Leu Pro Asn Asp Gly Asn His Trp	
46	10 15 20	
47	aac agg tta cga gat gac gca got aac tta aag agt aaa ggg att acc	367
48	Asn Arg Leu Arg Asp Asp Ala Ala Asn Leu Lys Ser Lys Gly Ile Thr	
49	25 30 35 40	
50	gct gtt tgg att cct cct gca tgg aag ggg act tcg caa aat gat gtt	415
51	Ala Val Trp Ile Pro Pro Ala Trp Lys Gly Thr Ser Gln Asn Asp Val	
52	45 50 55	
53	ggg tat ggt gcc tat gat ttg tac gat ctt ggt gag ttt aac caa aag	463
54	Gly Tyr Gly Ala Tyr Asp Leu Tyr Asp Leu Gly Glu Phe Asn Gln Lys	
55	60 65 70	
56	gga acc gtc cgt aca aaa tat ggc aca agg agt cag ttg caa ggt gcc	511
57	Gly Thr Val Arg Thr Lys Tyr Gly Thr Arg Ser Gln Leu Gln Gly Ala	
58	75 80 85	
59	gtg aca tct ttg aaa aat aac ggg att caa gtt tat ggg gat gtc gtg	559
60	Val Thr Ser Leu Lys Asn Asn Gly Ile Gln Val Tyr Gly Asp Val Val	
61	90 95 100	
62	atg aat cat aaa ggt gga gca gac ggg aca gag atg gta aat gcg gtg	607
63	Met Asn His Lys Gly Gly Ala Asp Gly Thr Glu Met Val Asn Ala Val	
64	105 110 115 120	
65	gaa gtg aac cga agc aac cga aac caa gaa ata tca ggt gaa tac acc	655
66	Glu Val Asn Arg Ser Asn Arg Asn Gln Glu Ile Ser Gly Glu Tyr Thr	
67	125 130 135	
68	att gaa gca tgg acg aaa ttt gat ttc cct gga aga gga aat acc cat	703
69	Ile Glu Ala Trp Thr Lys Phe Asp Phe Pro Gly Arg Gly Asn Thr His	
70	140 145 150	
71	tcc aac ttt aaa tgg cgc tgg tat cat ttt gat ggg aca gat tgg gat	751
72	Ser Asn Phe Lys Trp Arg Trp Tyr His Phe Asp Gly Thr Asp Trp Asp	
73	155 160 165	
74	cag tca cgt cag ctt cag aac aaa ata tat aaa ttc aga ggt acc gga	799
75	Gln Ser Arg Gln Leu Gln Asn Lys Ile Tyr Lys Phe Arg Gly Thr Gly	
76	170 175 180	
77	aag gca tgg gac tgg gaa gta gat ata gag aac ggc aac tat gat tac	847
78	Lys Ala Trp Asp Trp Glu Val Asp Ile Glu Asn Gly Asn Tyr Asp Tyr	
79	185 190 195 200	
80	ctt atg tat gca gac att gat atg gat cat cca gaa gta atc aat gaa	895
81	Leu Met Tyr Ala Asp Ile Asp Met Asp His Pro Glu Val Ile Asn Glu	
82	205 210 215	
83	ctt aga aat tgg gga gtt tgg tat aca aat aca ctt aat cta gat gga	943
84	Leu Arg Asn Trp Gly Val Trp Tyr Thr Asn Thr Leu Asn Leu Asp Gly	
85	220 225 230	
86	ttt aga atc gat gct gtg aaa cat att aaa tac agc tat acg aga gat	991
87	Phe Arg Ile Asp Ala Val Lys His Ile Lys Tyr Ser Tyr Thr Arg Asp	
88	235 240 245	
89	tgg cta aca cat gtg cgt aac acc aca ggt aaa cca atg ttt gca gtt	1039
90	Trp Leu Thr His Val Arg Asn Thr Thr Gly Lys Pro Met Phe Ala Val	
91	250 255 260	
92	gca gaa ttt tgg aaa aat gac ctt gct gca atc gaa aac tat tta aat	1087
93	Ala Glu Phe Trp Lys Asn Asp Leu Ala Ile Glu Asn Tyr Leu Asn	

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/798,278

DATE: 08/31/2004

TIME: 12:50:47

Input Set : N:\Crf3\RULE60\10798278.raw

Output Set: N:\CRF4\08312004\J798278.raw

94	265	270	275	280	
95	aaa aca agt tgg aat cac tcc gtg ttc gat gtt cct ctt cat tat aat				1135
96	Lys Thr Ser Trp Asn His Ser Val Phe Asp Val Pro Leu His Tyr Asn				
97	285	290	295		
98	ttg tac aat gca tct aat aat ggt ggc tat ttt gat atg aga aat att				1183
99	Leu Tyr Asn Ala Ser Asn Ser Gly Gly Tyr Phe Asp Met Arg Asn Ile				
100	300	305	310		
101	tta aat ggt tct gtc gta caa aaa cac cct ata cat gca gtc aca ttt				1231
102	Leu Asn Gly Ser Val Val Gln Lys His Pro Ile His Ala Val Thr Phe				
103	315	320	325		
104	gtt gat aac cat gac tct cag cca gga gaa gca ttg gaa tcc ttt gtt				1279
105	Val Asp Asn His Asp Ser Gln Pro Gly Glu Ala Leu Glu Ser Phe Val				
106	330	335	340		
107	caa tcg tgg ttc aaa cca ctg gca tat gca ttg att ctg aca agg gag				1327
108	Gln Ser Trp Phe Lys Pro Leu Ala Tyr Ala Leu Ile Leu Thr Arg Glu				
109	345	350	355	360	
110	caa ggt tac cct tcc gta ttt tac ggt gat tac tac ggt ata cca act				1375
111	Gln Gly Tyr Pro Ser Val Phe Tyr Gly Asp Tyr Tyr Gly Ile Pro Thr				
112	365	370	375		
113	cat ggt gtt cct tcg atg aaa tct aaa att gat cca ctt ctg cag gca				1423
114	His Gly Val Pro Ser Met Lys Ser Lys Ile Asp Pro Leu Leu Gln Ala				
115	380	385	390		
116	cgt caa acg tat gcc tac gga acc caa cat gat tat ttt gat cat cat				1471
117	Arg Gln Thr Tyr Ala Tyr Gly Thr Gln His Asp Tyr Phe Asp His His				
118	395	400	405		
119	gat att atc ggc tgg acg aga gaa ggg gac agc tcc cac cca aat tca				1519
120	Asp Ile Ile Gly Trp Thr Arg Glu Gly Asp Ser Ser His Pro Asn Ser				
121	410	415	420		
122	gga ctt gca act att atg tcc gat ggg cca ggg ggt aat aaa tgg atg				1567
123	Gly Leu Ala Thr Ile Met Ser Asp Gly Pro Gly Gly Asn Lys Trp Met				
124	425	430	435	440	
125	tat gtc ggg aaa cat aaa gct ggc caa gta tgg aga gat atc acc gga				1615
126	Tyr Val Gly Lys His Lys Ala Gly Gln Val Trp Arg Asp Ile Thr Gly				
127	445	450	455		
128	aat agg tct ggt acc gtc acc att aat gca gat ggt tgg ggg aat ttc				1663
129	Asn Arg Ser Gly Thr Val Thr Ile Asn Ala Asp Gly Trp Gly Asn Phe				
130	460	465	470		
131	act gta aac gga ggg gca gtt tcg gtt tgg gtg aag caa taaataagga				1712
132	Thr Val Asn Gly Gly Ala Val Ser Val Trp Val Lys Gln				
133	475	480	485		
134	acaagaggcg aaaattactt tcctacatgc agagcttcc gatcactcat acacccaata				1772
135	taaattggaa gctt				1786
137	<210> SEQ ID NO: 2				
138	<211> LENGTH: 516				
139	<212> TYPE: PRT				
140	<213> ORGANISM: Bacillus sp. KSM-AP1378				
141	<400> SEQUENCE: 2				
142	Met Lys Leu His Asn Arg Ile Ile Ser Val Leu Leu Thr Leu Leu				
143	-30	-25	-20		

RAW SEQUENCE LISTING DATE: 08/31/2004
PATENT APPLICATION: US/10/798,278 TIME: 12:50:47

Input Set : N:\Crf3\RULE60\10798278.raw
Output Set: N:\CRF4\08312004\J798278.raw

144 Ala Val Ala Val Leu Phe Pro Tyr Met Thr Glu Pro Ala Gln Ala His
145 -15 -10 -5 -1 1
146 His Asn Gly Thr Asn Gly Thr Met Met Gln Tyr Phe Glu Trp His Leu
147 5 10 15
148 Pro Asn Asp Gly Asn His Trp Asn Arg Leu Arg Asp Asp Ala Ala Asn
149 20 25 30
150 Leu Lys Ser Lys Gly Ile Thr Ala Val Trp Ile Pro Pro Ala Trp Lys
151 35 40 45
152 Gly Thr Ser Gln Asn Asp Val Gly Tyr Gly Ala Tyr Asp Leu Tyr Asp
153 50 55 60 65
154 Leu Gly Glu Phe Asn Gln Lys Gly Thr Val Arg Thr Lys Tyr Gly Thr
155 70 75 80
156 Arg Ser Gln' Leu Gln Gly Ala Val Thr Ser Leu Lys Asn Asn Gly Ile
157 85 90 95
158 Gln Val Tyr Gly Asp Val Val Met Asn His Lys Gly Gly Ala Asp Gly
159 100 105 110
160 Thr Glu Met Val Asn Ala Val Glu Val Asn Arg Ser Asn Arg Asn Gln
161 115 120 125
162 Glu Ile Ser Gly Glu Tyr Thr Ile Glu Ala Trp Thr Lys Phe Asp Phe
163 130 135 140 145
164 Pro Gly Arg Gly Asn Thr His Ser Asn Phe Lys Trp Arg Trp Tyr His
165 150 155 160
166 Phe Asp Gly Thr Asp Trp Asp Gln Ser Arg Gln Leu Gln Asn Lys Ile
167 165 170 175
168 Tyr Lys Phe Arg Gly Thr Gly Lys Ala Trp Asp Trp Glu Val Asp Ile
169 180 185 190
170 Glu Asn Gly Asn Tyr Asp Tyr Leu Met Tyr Ala Asp Ile Asp Met Asp
171 195 200 205
172 His Pro Glu Val Ile Asn Glu Leu Arg Asn Trp Gly Val Trp Tyr Thr
173 210 215 220 225
174 Asn Thr Leu Asn Leu Asp Gly Phe Arg Ile Asp Ala Val Lys His Ile
175 230 235 240
176 Lys Tyr Ser Tyr Thr Arg Asp Trp Leu Thr His Val Arg Asn Thr Thr
177 245 250 255
178 Gly Lys Pro Met Phe Ala Val Ala Glu Phe Trp Lys Asn Asp Leu Ala
179 260 265 270
180 Ala Ile Glu Asn Tyr Leu Asn Lys Thr Ser Trp Asn His Ser Val Phe
181 275 280 285
182 Asp Val Pro Leu His Tyr Asn Leu Tyr Asn Ala Ser Asn Ser Gly Gly
183 290 295 300 305
184 Tyr Phe Asp Met Arg Asn Ile Leu Asn Gly Ser Val Val Gln Lys His
185 310 315 320
186 Pro Ile His Ala Val Thr Phe Val Asp Asn His Asp Ser Gln Pro Gly
187 325 330 335
188 Glu Ala Leu Glu Ser Phe Val Gln Ser Trp Phe Lys Pro Leu Ala Tyr
189 340 345 350
190 Ala Leu Ile Leu Thr Arg Glu Gln Gly Tyr Pro Ser Val Phe Tyr Gly
191 355 360 365
192 Asp Tyr Tyr Gly Ile Pro Thr His Gly Val Pro Ser Met Lys Ser Lys

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/798,278

DATE: 08/31/2004
TIME: 12:50:47

Input Set : N:\Crf3\RULE60\10798278.raw
Output Set: N:\CRF4\08312004\J798278.raw

193	370	375	380	385
194	Ile Asp Pro Leu Leu Gln Ala Arg Gln Thr Tyr Ala Tyr Gly Thr Gln			
195	390	395	400	
196	His Asp Tyr Phe Asp His His Asp Ile Ile Gly Trp Thr Arg Glu Gly			
197	405	410	415	
198	Asp Ser Ser His Pro Asn Ser Gly Leu Ala Thr Ile Met Ser Asp Gly			
199	420	425	430	
200	Pro Gly Gly Asn Lys Trp Met Tyr Val Gly Lys His Lys Ala Gly Gln			
201	435	440	445	
202	Val Trp Arg Asp Ile Thr Gly Asn Arg Ser Gly Thr Val Thr Ile Asn			
203	450	455	460	465
204	Ala Asp Gly Trp Gly Asn Phe Thr Val Asn Gly Gly Ala Val Ser Val			
205	470	475	480	
206	Trp Val Lys Gln			
207	485			

209 <210> SEQ ID NO: 3

210 <211> LENGTH: 1753

211 <212> TYPE: DNA

212 <213> ORGANISM: Bacillus sp. KSM-K38

213 <220> FEATURE:

214 <221> NAME/KEY: sig_peptide

215 <222> LOCATION: (162)..(224)

216 <223> OTHER INFORMATION:

217 <221> NAME/KEY: mat_peptide

218 <222> LOCATION: (225)..()

219 <223> OTHER INFORMATION:

--> 220 <221> CDS

221 <222> LOCATION: (162)..(1664)

222 <223> OTHER INFORMATION:

--> 223 <400> 3

224	gtatgcgaaa cgatgcgcaa aactgcgcaa ctactagcac tcttcaggga ctaaaccacc	60
225	tttttccaa aaatgacatc atataaacaa atttgtctac caatcactat ttAAAGCTGT	120
226	ttatgatata tgtaagcggtt atcattaaaa ggaggtatTTT g ATG AGA AGA TGG GTA	176
227	Met Arg Arg Trp Val	
228	-20	
229	gta gca atg ttg gca gtg tta ttt cct tcg gta gta gtt gca	224
230	Val Ala Met Leu Ala Val Leu Phe Leu Phe Pro Ser Val Val Ala	
231	-15 -10 -5 -1	
232	gat gga ttg aac ggt acg atg cag tat tat gag tgg cat ttg gaa	272
233	Asp Gly Leu Asn Gly Thr Met Met Gln Tyr Tyr Glu Trp His Leu Glu	
234	1 5 10 15	
235	aac gac ggg cag cat tgg aat cgg ttg cac gat gat gcc gca gct ttg	320
236	Asn Asp Gly Gln His Trp Asn Arg Leu His Asp Asp Ala Ala Leu	
237	20 25 30	
238	agt gat gct ggt att aca gct att tgg att ccg cca gcc tac aaa ggt	368
239	Ser Asp Ala Gly Ile Thr Ala Ile Trp Ile Pro Pro Ala Tyr Lys Gly	
240	35 40 45	
241	aat agt cag gcg gat gtt ggg tac ggt gca tac gat ctt tat gat tta	416
242	Asn Ser Gln Ala Asp Val Gly Tyr Ala Tyr Asp Leu Tyr Asp Leu	

RAW SEQUENCE LISTING ERROR SUMMARY DATE: 08/31/2004
PATENT APPLICATION: US/10/798,278 TIME: 12:50:48

Input Set : N:\Crf3\RULE60\10798278.raw
Output Set: N:\CRF4\08312004\J798278.raw

Please Note:

use of n and/or Xaa have been detected in the Sequence Listing. Please review the
Sequence Listing to ensure that a corresponding explanation is presented in the <220>
o <223> fields of each sequence which presents at least one n or Xaa.

eq#:33; N Pos. 16,17,18
eq#:34; N Pos. 16,17,18
eq#:35; N Pos. 16,17,18
eq#:36; N Pos. 16,17,18
eq#:42; N Pos. 27
eq#:48; N Pos. 19,20,21

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/798,278

DATE: 08/31/2004

TIME: 12:50:48

Input Set : N:\Crf3\RULE60\10798278.raw

Output Set: N:\CRF4\08312004\J798278.raw

L:29 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:29 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:1,Line#:0
L:32 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:1
L:32 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:1,Line#:0
L:32 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:1,Line#:0
L:220 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:3
L:220 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:3,Line#:0
L:223 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:3
L:223 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:3,Line#:0
L:223 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:3,Line#:0
L:653 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!
L:656 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:33
L:657 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:33 after pos.:0
L:665 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!
L:668 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:34
L:669 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:34 after pos.:0
L:677 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!
L:680 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:35
L:681 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:35 after pos.:0
L:689 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!
L:692 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:36
L:693 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:36 after pos.:0
L:746 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!
L:749 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:42
L:750 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:42 after pos.:0
L:803 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!
L:806 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:48
L:807 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:48 after pos.:0